

Recommender Systems

JJ Vie

January 21, 2020

Collaborative filtering



Sacha	?	5	2	?
Ondine	4	1	?	5
Pierre	3	3	1	4
Joëlle	5	?	2	?

Collaborative filtering



Sacha	3	5	2	2
Ondine	4	1	4	5
Pierre	3	3	1	4
Joëlle	5	2	2	5

Algorithm: K -nearest neighbors

To recommend movies:

- ▶ We choose a **similarity score** between people
- ▶ We find **K nearest neighbors** of someone
- ▶ We recommend to this person what they liked that this person did not watch

Our data

	007	Batman 1	Shrek 2	Toy Story 3	Star Wars 4	Twilight 5
Alice	+	—	0	+	0	—
Bob	—	0	+	—	+	+
Charles	+	+	+	+	—	—
Daisy	+	+	0	0	+	—
Everett	+	—	+	+	—	0

What similarity score should we choose?

Computing the score

	007	Batman 1	Shrek 2	Toy Story 3	Star Wars 4	Twilight 5
Alice	+	−	0	+	0	−
Charles	+	+	+	+	−	−
Score	+1	−1		+1		+1

$$\text{score}(\text{Alice}, \text{Charles}) = 3 + (-1) = 2$$

	007	Batman 1	Shrek 2	Toy Story 3	Star Wars 4	Twilight 5
Alice	+	−	0	+	0	−
Bob	−	0	+	−	+	+
Score	−1			−1		−1

$$\text{score}(\text{Alice}, \text{Bob}) = -3$$

Alice is closer to Charles than Bob

Similarity score between people

	Alice	Bob	Charles	Daisy	JJ
Alice	4	-3	2	1	3
Bob	-3	5	-3	-1	-2
Charles	2	-3	6	2	3
Daisy	1	-1	2	4	-1
Everett	3	-2	3	-1	5

Who are the 2 nearest neighbors from Alice?

Computing the predictions

	007	Batman 1	Shrek 2	Toy Story 3	Star Wars 4	Twilight 5
Alice	+	—	?	+	?	—
Charles	+	+	+	+	—	—
Daisy	+	+	0	0	+	—
Everett	+	—	+	+	—	0

Knowing her neighbors, how likely will Alice enjoy these movies?

Computing the predictions

	007	Batman 1	Shrek 2	Toy Story 3	Star Wars 4	Twilight 5
Alice	+	−	+	+	−	−
Charles	+	+	+	+	−	−
Daisy	+	+	0	0	+	−
Everett	+	−	+	+	−	0

We can compute the average: $\text{prediction}(\text{Alice}, \text{Star Wars 4}) = -0,333\dots$

Let's code!

- ▶ `compute_score(i, j)`
- ▶ `compute_all_scores()`
- ▶ `nearest_neighbors(i)`
- ▶ `compute_prediction(i, i_{film})`
- ▶ `compute_all_predictions(i)`